# 5

# Pragmatic and Use Designs

Kellie Frost

In the previous chapter, we set out a 'postpositivist and methods' branch of program evaluation to situate quality assurance in language programs. We examined the work of quality assurance agencies throughout the global English language teaching and learning industry. Defining and enhancing quality, we saw, informs potential students when choosing a school as well as being integral to the promotion of national reputation in the field. As shown in the work of NEAS, we argued that this design was perhaps the most appropriate for commercial English Language Teaching (ELT) schools. To illustrate the design, we presented an example of how the quality assurance design was conducted in a hybrid language school.

In this chapter, we move on to discuss design approaches that can be clustered under the pragmatic paradigm and use branch (Mertens &

K. Frost (✉)
School of Languages and Linguistics, Faculty of Arts,
The University of Melbourne, Melbourne, VIC, Australia
e-mail: kmfrost@unimelb.edu.au

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

P. Gruba, *Designs for Language Program Evaluation*,
https://doi.org/10.1007/978-3-031-68926-0_5

Wilson, 2019), with a particular focus on *Developmental Evaluation* (Patton, 2011, 2015a). Developmental evaluation falls within the wider domain of utilization-focused evaluation (UFE), also developed by Patton (2008). UFE, as an approach, is underpinned by a principle of usefulness to intended users; evaluation processes and outcomes should provide intended users—those with the authority and responsibility to act to ensure and promote program quality—with information to facilitate effective, evidence-based decision making and action. As Patton (2008, 2011, 2015a) explains, to enable usefulness, evaluators must identify intended users at the outset of an evaluation project, and ensure that these people are actively involved in the evaluation process from beginning to end; intended users should understand and feel a sense of ownership of the objectives of the evaluation process, the measures used to define program quality and success, and the establishment and enactment of a plan for translating evaluation findings into changes/innovations in the program. The evaluator must also establish and monitor shared understandings, as well as ongoing participation and investment over the life of an evaluation project. This approach represents a departure from the quality assurance focus within the postpositivist paradigm and methods branch that measure program success/quality against a predetermined set of criteria; arguably, the processes of quality assurance designs function primarily to meet external accountability and regulatory requirements.

Developmental evaluation, as an example of UFE, remains tied to this principle of usefulness to intended users, but in this case, the approach is specifically tailored to highly complex and dynamic contexts, within which the ongoing development of *innovations* is a necessary and central aspect. A developmental evaluation is suitable, for example, in contexts in which program leaders and other stakeholders must engage in innovative design to adapt to shifting conditions and/or an emerging awareness of a complex context, with an evolving set of stakeholder needs, constraints and affordances, which impact the efficacy of program design and implementation. As set out in this chapter, we feel this approach to evaluation is well suited to addressing concerns and needs arising from a recent and rapid growth of English Medium Instruction (EMI) in universities throughout Asia. EMI can be broadly defined as "the use of the English language to teach academic subjects (other than English itself) in

countries or jurisdictions where the first language of the majority of the population is not English" (Macaro et al., 2018, p. 37). Over the past decade, higher education institutions across numerous countries in Asia, including but not limited to China, Japan, Vietnam, and more recently, Indonesia, have introduced EMI and/or expanded the number of courses they deliver in English. This trend has been particularly pronounced among elite universities, with EMI serving as a marker of prestige and global relevance, a reflection of an increasing emphasis on the importance of internationalization as a national policy agenda in numerous countries (Fenton-Smith et al., 2017; Lamb et al., 2021). Simultaneously, national, and provincial education ministries have been simultaneously calling upon universities to deliver improvements in the English skills of their graduating students, as it has become clear that expected benefits of EMI on language proficiency have not always materialized (Macaro, 2022).

A growing body of research examining this phenomenon has highlighted the variegated and contingent nature of EMI programs and their implementation, and raised numerous concerns over program design, resourcing, and the language proficiency needs of both teaching staff and students, to name but a few. Nonetheless, EMI in tertiary education in Asia is by now widely viewed as an "unstoppable train" (Macaro, 2015), with disciplinary program leaders, lecturers, and English teachers compelled to flexibly adapt not only their curricula, but also their pedagogic and linguistic strategies, goals and practices, to grapple with the demands and challenges of delivering disciplinary content in English, a language which is for many staff and students alike, the second or even third or fourth in their repertoires (Fenton-Smith et al., 2017).

With this in mind, we present EMI as a phenomenon which demands a view of innovation and evaluation as a set of inextricably linked and mutually constituting processes. Such a view, we argue, is necessary to enable continual improvements to program quality in the context of a phenomenon which has developed with such rapidity and complexity that a clear characterization of best practice remains elusive. We further stress a need to recognize and, indeed, utilize evaluation as a means of developing language programs that are innovative in ways that align with social justice principles of inclusion and decolonization (e.g. Kubota, 2022), promote sustainable and culturally meaningful engagement in teaching and learning activities, and which enable knowledge building

across both language and disciplinary areas. As always, we argue that evaluation can play a crucial role in fostering success in language programs. In the context of EMI, given its expanding scale and potential societal consequences across different countries and communities in Asia, this is underscored by a sense of urgency that is perhaps unprecedented.

The chapter is organized as follows. We begin as we have in previous chapters, by first outlining the design theory behind the pragmatic paradigm and use branch of evaluation. We then situate the design in the context of language program evaluation by summarizing synergies between this branch of evaluation and prominent approaches to language program design in our field of applied linguistics. This is followed by a description of developmental evaluation as the approach we consider most apt for addressing questions of EMI program design and success, particularly in the context of higher education in Asia. After a review of current literature on EMI in Asia, we provide an example of an EMI program in the Faculty of Social and Political Sciences of a prominent Indonesian university.

## Design Theory

As in the previous chapter, we again adopt Mertens and Wilson's (2019) framework to situate utilization-focused evaluation designs within the pragmatic paradigm and use branch as shown in Table 5.1.

A central tenet of the pragmatic paradigm and use branch, which includes but is not limited to developmental evaluation nor to UFE more broadly, is that evaluation processes and findings must be useful to primary stakeholders, informing their decisions and driving actions in relation to policy and/or program design and improvements (Mertens & Wilson, 2019). Consequently, in comparison to postpositivist and methods approaches, within pragmatic and use-based designs, the importance of establishing truth claims and an objective definition of a singular reality is de-emphasized; instead, the assumption is that even if a single reality exists, different people will have their own unique interpretation of it. Thus, rather than generating objective findings in relation to external measures of program quality, an evaluator's concern is instead to understand stakeholder perspectives, values, and goals. As a result, ontological

**Table 5.1** Summary of the pragmatic paradigm and use branch

| Area | Description | Related references in applied linguistics |
|---|---|---|
| Description (What is it?) | Common sense, practical thinking informs the scientific discovery of truth | Lancaster (2018) |
| Axiological assumption (What is its value?) | Knowledge is pursued to achieve ends, and is influenced by evaluator values and politics | Lynch (2003) |
| Ontological assumption (What is it? Or … What *could* it be?) | Though reality is singular, each person has a unique interpretation | Phan and Doan (2020) |
| Epistemological assumption (How do we produce conceptual knowledge?) | Relationships in evaluation are gauged by the evaluator on their appropriateness to the work at hand | Mirhosseini (2018a) |
| Methodological assumption (How is data collected?) | Choice of methods is driven by questions and aims, and mixed methods can be used flexibly | Bolton et al. (2023) |

Based on Mertens and Wilson (2019, p. 87)

questions are somewhat sidestepped, to enable a focus on practical thinking and relevance to the needs and goals of intended users. In terms of epistemological assumptions, relationships with different stakeholders are seen to provide the evaluator with the means of knowing the perspectives, values, and goals relevant to the aims of the program and the evaluation process.

Ethical (axiological) considerations are also largely driven by the values of program decision makers, and as already mentioned, it is the role of the evaluator to ensure participation on the part of those who are intended as the primary users and beneficiaries of the evaluation process and findings. Finally, the appropriateness of methods (qualitative or quantitative) can only be determined in light of the focus and aims of the evaluation, as determined through engagement with intended users and other stakeholders. This enables flexible and mixed methods designs (methodological assumption), which is also a distinguishing feature of pragmatic and use designs compared to postpositivist and methods designs, where, as set out in the previous chapter, quantitative methods, as a means of promoting objectivity and ensuring reliability of findings, are typically preferred.

Within the pragmatic paradigm and use branch, we can thus conceptualize the value of an evaluation in terms of its practical utility and the beneficial consequences it generates for intended users within the program in question. As Norris (2016) explains, in educational contexts, this focus emerged in the mid-1960s in response to realizations that leaders of poorly performing schools "required richer information that would lead to improvements in practices, not merely the 'finding' that they were underperforming" (p. 170). Mertens and Wilson (2019) similarly describe pragmatic and use approaches to evaluation in terms of a need to move beyond "evaluation for the sake of evaluation" (p. 86). To summarize then, the purpose of evaluation within this branch is no longer primarily to generate objective measures of program quality and/or impacts—a key tenet of postpositivist approaches described in the previous chapter. Instead, evaluations within the pragmatic paradigm and use branch are intended to generate insights into the potential consequences and implications of alternative decision options, whether these be in relation to program development, implementation, and/or program changes. Thus, the idea is that evaluation entails providing information to primary stakeholders (the intended users of evaluation findings) to inform effective planning and decision making, depending on their own concerns, needs, and objectives.

In the remainder of this chapter, we focus specifically on developmental evaluation. As outlined at the beginning of the chapter, this is an approach oriented to enabling and supporting innovative practices, which here we situate in the context of newly emerging EMI programs in tertiary education in Asia. There are alternative approaches to evaluation within the pragmatic paradigm and use branch, which similarly emphasize relevance and usefulness for informing stakeholder decision making and action, and which would be appropriate for evaluations of more well-established and stable language programs. As set out by Mertens and Wilson (2019), these include the "context, input, process, product" (CIPP) model, the model of practical participatory evaluation (in which program stakeholders are viewed as equal partners in the evaluation process), and the already mentioned utilization-focused evaluation model, which encompasses but is not limited to developmental evaluation. An account of each of these is beyond the scope of the current chapter but

see Mertens and Wilson (2019, pp. 93–110) for details. We shift attention now to developmental evaluation, which is the approach we deem most suitable to evaluations of language programs in contexts characterized by dynamicity and complexity, as we argue is the case for EMI programs in Asia, particularly in countries where the phenomenon has only recently taken hold.

## Situating the Design

As part of our efforts to situate program evaluation within the field of applied linguistics, and in relation to language teaching and learning programs specifically, it is worth noting at the outset that much, if not all, of the language program evaluation work to date can be characterized within a pragmatic paradigm (see for example, Elder, 2009; Lynch, 1996, 2019; Norris, 2009, 2016). The pragmatic paradigm and use branch of program evaluation, in emphasizing practical, decision-making relevance and beneficial consequences for stakeholders, also aligns well with needs analysis-based approaches to language curriculum design, including in areas of language for specific purposes (LSP) (e.g. Basturkmen, 2018; Flowerdew, 2012), and task-based language teaching (TBLT) (Ellis et al., 2019). In these types of programs, learners, their needs, and the linguistic demands they face in relevant domains (e.g. higher education, or specific domains of work and/or social life) are at the centre of program design, including the specification of learning objectives, the development of teaching and learning tasks, and the design of assessment tools and/or formal tests to evaluate learning outcomes (e.g. Smith et al., 2022). However, as we have mentioned in earlier chapters, the interface between research in second language acquisition, and language teaching programs and practices, is ambiguous, at best (Chapelle, 2021; Norris, 2016). Thus, it remains difficult to establish a priori a theory of change to support language program evaluation.

A lack of unequivocal 'theory of change' adds to the challenge of developing EMI programs, especially where development and implementation are occurring rapidly in dynamic settings, as is the case in higher education in Asia, and further speaks to the relevance of developmental

evaluation to these contexts. As we have already set out, we situate developmental evaluation as a productive approach to enhancing both potential for innovation as well as program quality in EMI contexts. Before discussing this further, we first provide a description of the eight guiding principles behind developmental evaluation in Table 5.2.

The eight principles set out in Table 5.2 serve the purpose of addressing two initial questions that should be asked before embarking on a program evaluation: (1) which approach is best suited to the context and goals of the evaluation—if there is a developmental purpose, a utilization focus and an innovation niche, then developmental evaluation is likely to be a suitable approach, (2) will it be possible to adopt systems thinking and a complexity perspective in a meaningful way (for both the evaluator and the primary stakeholders/intended users), and (3) are co-creation

**Table 5.2** Principles of developmental evaluation

| Principle | Description |
|---|---|
| Developmental purpose | Illuminate, inform, and support innovation & change development/ theory of change elaboration |
| Evaluation rigor | Ask probing questions, promote evaluative thinking, use appropriate methods, stay empirically grounded |
| Utilization focus | Involve intended users from beginning to end of evaluation process, in evaluative thinking (above) and ongoing use of evaluation insights to inform innovation/adaptation development (below) |
| Innovation niche | Identify the nature of innovation/adaptation within broader processes of change; identify complex 'problem' that innovation is designed to address |
| Complexity perspective | Expect plans, goals, targets to evolve as innovation develops; expect and monitor for unintended consequences; evaluation design should be emergent and adaptive |
| Systems thinking | Attend to interrelationships, different perspectives, boundaries and how these are and could be understood and constructed/re-constructed |
| Co-creation | Innovation and evaluation are inextricably linked and feed into each other iteratively over time to enact processes of change |
| Timely feedback | Feedback from the evaluation process is delivered as it emerges in an ongoing process to inform innovation and adaptation development |

Based on Patton (2015b, pp. 291–308)

and timely feedback feasible (in terms of innovation design, resourcing and also relationships relevant to the evaluation)—as shown in the table above, these are fundamental tenets of this approach to evaluation. In considering these questions, we turn to thinking further about this notion of innovation, and how it might specifically apply to language program evaluation, including the somewhat problematic aspect of a theory of change.

As we now know, developmental evaluation, like other utilization-focused approaches, is underpinned by a focus on usefulness, and is designed to address decision-making needs in programs characterized by emerging innovations. As part of this, there is an impetus to provide, through the evaluation process, ongoing feedback to intended users and associated stakeholders as they engage in iterative program design and implementation, to inform and improve innovations and thus to enhance overall program quality. Patton (2015a) further explains that developmental evaluation is used to track and inform innovation and systems change in complex systems. In the context of language program evaluation, we can think of system change in two ways. Firstly, in terms of the dynamics of the educational contexts within which EMI programs are emerging; as described in the section below, EMI in its various manifestations has emerged in the context of other education and social system changes as part of wider processes of globalization and a perceived need for internationalization, with little yet known about how to systematically address the range of complex and highly situated challenges that this brings. These challenges have already prompted the emergence of various innovations in program design as well as in teaching and learning practices, albeit with little, if any, planning. Secondly, both language and language learners can be thought of as interconnected complex systems within these larger social systems (Larsen-Freeman & Cameron, 2008, see also Chap. 2).

For the purposes of language program evaluation, developmental evaluation designs can not only be used to inform overall program development and teaching innovations, but also to foster innovations in second language acquisition theories (including, for example, iterative evaluations of their appropriateness and explanatory power) in ways that strengthen links between language learning processes, teaching practices,

and program design. Looking ahead, developmental designs can also inform a theory of change that can inform EMI program evaluations in diverse settings. Before illustrating developmental evaluation design in the context of EMI in Indonesia, we turn now to a characterization of EMI programs, with a focus on Asia more broadly, to further draw out the potential innovation niche, not to mention the complexity of EMI and the potential for developmental evaluation to advance the opportunities and mitigate the challenges this complexity brings.

## Key Characteristics of EMI Programs and Prominent Research

As noted, there has been a widespread expansion of EMI programs introduced in higher education contexts across Asia in recent years. Macaro (2022, pp. 535–536), drawing on Richards and Pun (2021), identifies at least four different models of EMI in higher education: (1) the preparatory year model, in which students are provided with intensive English language support for a specified period of up to a year; (2) the pre-institutional selection model, where access to EMI programs is regulated through English proficiency requirements at transitions from secondary into tertiary education, and/or from undergraduate to postgraduate study; (3) the institutional concurrent support model, in which access depends primarily on secondary school performance in relevant disciplinary subjects, and ongoing English language support is available as part of higher education; (4) the multilingual model, which can align with any of the three previous models, but includes flexibility in the language used for teaching content subjects.

National policies to implement and expand EMI in higher education in Asia are widely viewed as being driven by global imperatives, including the status of English as an international language, a perceived need for internationalization in higher education, and associated competition in both the recruitment of international students and the production and marketability of local students with international competencies (e.g. Fenton-Smith et al., 2017; Macaro, 2022; Macaro et al., 2018). Across

the literature, two main assumptions are widely identified as underlying EMI policy drivers: (1) EMI will result in higher levels of English proficiency among graduates, and (2) EMI will not negatively impact disciplinary learning/standards of achievement (e.g. Macaro et al., 2018; Shohamy, 2012). However, as Macaro et al. (2018) emphasize, there is a dearth of research to date examining the effects of EMI on language learning, and a dearth of research examining effects on content learning. This, they note, means that "any cost-benefit evaluation of EMI is inconclusive at best and impossible at worst" (p. 64). There thus remains an urgent need to evaluate the impacts of a shift to EMI, to determine whether the educational objectives behind the policy imperative are achievable. More importantly, perhaps, there is a need to evaluate the social justice implications of this increasing emphasis on English in terms of longer-term societal impacts, such as the potential loss of local language varieties, and increasing social inequalities, as discussed further below.

Not surprisingly, given the diverse range of global imperatives behind the trend, the implementation of EMI is not necessarily driven by deliberate strategies or an explicit language program design, and may instead be an unintended consequence of other policy decisions at the supranational level (Macaro et al., 2019). Referring to a recent upsurge in the number of EMI programs in universities across Europe, where the phenomenon has a much longer history, Macaro et al. (2019) point to the increasing importance of performance indicators, including global university rankings, as measures of competitiveness in an era of knowledge-based economies, human capital, and internationalization. At the level of individual universities, the development and implementation of EMI programs are thus not necessarily concerned with language per se and may not encompass any explicit language learning objectives. In the context of university education across Asia, where the rapid expansion of EMI is a more recent phenomenon, Macaro et al. (2019) point out that "many of these EMI programmes have been established without any real planning or thought for the potential implications" (p. 237). This is consistent with Fenton-Smith et al. (2017), who report that many institutions in Asia introduced EMI programs without any real stakeholder consultation, and in the absence of any coherent language policy or

design goals. Macaro et al. (2018), based on a systematic review of 83 studies of EMI in higher education across various different countries, also found little evidence across the literature of any substantive input from teachers or students in EMI design and implementation, in Asia or elsewhere.

Given the absence of a coherent policy and general lack of stakeholder consultation to date, it is also not surprising that there is little consistency in how EMI is designed and implemented across different contexts, with approaches to pedagogy and the extent to which language and content-related learning objectives coexist, varying widely. Macaro et al. (2018), for example, in a survey of policies and practices across numerous contexts in Europe and Asia, found that the pedagogical approach adopted (when rarely specified), depended not only on the particular institutional policy but also varied between subject areas, between teachers and classrooms in the same discipline in the same institution, and between classes delivered by a single teacher over time. Even in institutions with a policy of English only as the language of instruction, teachers use a combination of English and their first language, depending on a range of factors, including their own sense of legitimacy/proficiency in English, as well as their perception of students' ability to understand content in English. Other practices also vary, including levels of interaction and the speed of delivery of content, depending on whether English or the mother tongue was used as the primary language in classrooms.

Wider concerns have also already been raised about the social justice implications of EMI, including the potential to entrench existing structural inequalities in societies (Lamb et al., 2021; Macaro et al., 2018; Shohamy, 2012). As Shohamy (2012) reminds us, EMI at university level cannot be detached from the wider ideologically context, through which certain knowledges and ways of knowing are valued over others, and certain language practices and associated identities are legitimized as belonging in a society, while others are de-valued and marginalized. For minority language speakers, where neither the national language nor English is their first language, Shohamy

emphasizes the potential for a shift to EMI to compound their marginalization, given that disciplinary learning outcomes are typically lower in a third language compared to a second language. If we consider this in the context of a shift to EMI in higher education in Asia, where a range of intersecting policy measures have been introduced to add instrumental value to English, including incentives for university staff to publish in English language academic journals, including links to promotion and in some cases, remuneration, as well as requirements for graduate students to publish in English in order to be awarded Master or Doctorate degrees, we can see that the potential for exclusion is amplified. Debates around EMI have also focused on the question of which variety of English should be promoted, amidst concerns about the appropriateness and relevance of nativized varieties in multilingual contexts where English functions primarily as a lingua franca. Such concerns highlight the potential negative impact of what has been termed "Englishisation" (Hultgren, 2014) on students' discipline-specific proficiency in local languages.

In this context of contested values, evaluation is urgently needed to understand not only the impacts of different approaches to EMI on the learning and other lived experiences of students, but also to support the development of new innovations, such as training resources to support program staff as they traverse domains of disciplinary and language teaching and learning. There is also a need for new assessment tools to measure lecturer readiness for EMI, the readiness of teaching materials (accessibility of oral/written/multimodal texts and potential for knowledge building), and the readiness of students to learn in and through the medium of English. Developmental evaluation offers an approach well suited to the context of EMI in Asia, given the rapid pace of change and the intended scale and scope of program intentions, and especially given that programs are being simultaneously innovated and implemented, sometimes, as already noted, with very little pre-planning and deliberate design. EMI program development and implementation in Asian countries thus represent Patton's (2015b) notion of *ongoing adaptive development*.

## Illustrating the Design: EMI in an Indonesian University

Moving from theory to practice, we turn now to an illustrative example of developmental evaluation applied to an EMI program in an Indonesian university. In constructing a case study for evaluation for this chapter, we were informed by a growing body of research examining EMI in higher education which has emerged over the past decade, including edited volumes dedicated to the topic in the Asian context and elsewhere (e.g. Fenton-Smith et al., 2017; McKinley & Galloway, 2022), as well as an increasing number of studies focused on EMI policies, pedagogies, and impacts on teachers, students, and learning outcomes in various national contexts in Asia, including Vietnam (Nguyen, 2022), China (Zhao & Dixon, 2017), Korea (Kim, 2017), Japan (Hino, 2017), and more recently, Indonesia (Dewi, 2017; Lamb et al., 2021), where the phenomenon has most recently taken hold. Where these studies have focused on a particular education program within a particular institution, the aims and concerns that researchers are seeking to address are often, if not always, relevant to language program evaluation. However, these studies are rarely framed as evaluations, per se, and are often aimed at addressing a single aspect or limited number of aspects of an EMI program.

What exists, as a result, is a fragmented understanding of EMI and the factors behind the success or otherwise of this kind of program. As we have advocated throughout this book, a systematic evaluation framework provides a means of understanding and articulating the objectives underlying a language program. Developmental evaluation, as we have already argued, is particularly well suited to supporting innovation in the complex and dynamic settings which characterize EMI program development and implementation. This is especially the case in the Indonesian higher education context, where EMI is in its early stages compared to other countries in Asia (Lamb et al., 2021). We will again use the BetterEvaluation Rainbow Framework to illustrate how a developmental evaluation design of an EMI program might be implemented in this context. Bearing in mind the eight key principles of developmental evaluation shown earlier in Table 5.2, we will aim to highlight those which distinguish

developmental evaluation from other utilization-focused approaches, particularly the *innovation niche*, and a *complexity perspective* and *systems thinking*, in our illustration, below.

Leading universities in Indonesia have well-established international programs in many discipline areas, which have long been delivered in English, but these have strict English proficiency entry requirements and thus are generally accessible only by the small proportion of students who have a strong background in English, such as those who have completed primary and/or secondary schooling in international schools, for example (Zein et al., 2020). However, English has become a highly valued skill for employment, even in positions that are based and operate mainly within Indonesia. As a result, universities are now under pressure to develop the English communication skills of all graduates, and as a result, efforts are being made to introduce EMI courses into regular degree programs, where students have previously been taught and assessed solely in Bahasa Indonesia.

In such contexts, several challenges have so far emerged (see Dewi, 2017; Lamb et al., 2021), similar to those reported in other EMI contexts in Asia, which call for a range of innovations not only in terms of curriculum design, teacher training, and student assessment, but also in terms of how English is conceptualized and potentially integrated into teaching practices alongside Bahasa Indonesia and the local languages of students and teaching staff. As we know, educational outcomes are typically lower for students who are taught and assessed in a second language compared to those taught in their mother tongue, and outcomes worsen for those learning in a third or further additional language (Shohamy, 2012). Reflecting a *complexity perspective*, a key underlying evaluation question is thus as follows: How can we help mitigate the risk of EMI functioning to exacerbate social inequalities that already function to privilege those with high levels of English, supporting an EMI program to achieve learning objectives and improve employment opportunities for a wider proportion of graduates, regardless of students' previous access to English language education?

We situate our design through a constructed example of an EMI program of an M.A. in political science at an Indonesian university. We note here that this example is derived from an amalgam of our experiences in

working with programs in different faculties and across different disciplines in a number of Indonesian universities, where the focus has been on developing and evaluating new English teaching and assessment materials as part of teaching collaborations between our institution and partner Indonesian institutions. While our discussion is framed around a single program, this is intended as a heuristic—we are not providing details here of an actual evaluation conducted on a particular program.

## Manage

As we have seen, the pragmatic paradigm and use branch emphasizes the need to ensure that evaluation processes and findings are useful to and used by intended stakeholders—those with the authority and responsibility to manage and enhance program design, implementation, and ongoing quality of delivery. We were initially invited by the Faculty of Social and Political Sciences at a university in Indonesia to share our expertise in English language teaching and assessment with their academic skill support team, to assist them in adapting the existing curriculum of one of the compulsory subjects in the M.A. for EMI; they were envisaging a staged move to EMI in their regular program, beginning with a pilot subject, which we agreed was a sensible approach. They were originally seeking advice as to how they might provide scaffolding with reading tasks to support students with lower levels of English proficiency, as well as how they might encourage these students, who were perceived to be very reluctant to use English in oral communication, to develop interactional skills and oral presentation skills in English. They also asked us which of the available English language tests they should use to assess learning outcomes from their new EMI courses. Given the diversity of their students, many of whom had minority language backgrounds not to mention vastly different levels of experiences in learning English, we immediately recognized the complexity behind their request for assistance, and moved quickly to discussions of a wider mapping of the program objectives, affordances, and constraints (especially in terms of institutional resourcing and available expertise/training requirements), and potential ways to go about developing appropriate pedagogic

materials and assessment tools that would be both sustainable and reflective of the needs and goals of their students.

Our concerns led to a series of meetings with senior management in the university and faculty, to gain an understanding of the perspectives and key concerns of primary decision makers, whose support would be critical. These included the University Vice Rector for Academic and Student Affairs, since this person had an overview of EMI policy and programs across the university, as well as an understanding of the rationale behind the introduction of EMI in this institutional context. We also met with Faculty of Social and Political Sciences leaders, including the Vice Dean of Academic and Student Affairs, responsible for overseeing the different degree programs and courses within the faculty, including the EMI programs, and the Head of the M.A. who would oversee the specific adaptations to the regular program as part of a staged move to EMI. There was a strong motivation among senior management and program leadership to enhance the prestige of the program by evidencing English proficiency gains for students, which they saw as necessary for positioning their graduates for internationally oriented employment in high-profile organizations in Indonesia. We agreed that given the complexity involved in addressing the needs of their diverse student cohort, a trial-and-error approach would be necessary, at least to some extent, and the timeframe for a concurrent development-evaluation process would span at least two iterations of the delivery of the first 'trial' EMI course, so a period of at least one year. Resourcing for our involvement as an external evaluation team was limited, and so an important part of planning involved surveying the expertise of the program staff in Indonesia to identify gaps and to determine the nature and timing of our involvement in the development of teaching and assessment materials, as well as in evaluations of quality, to ensure our contribution would be most productive.

## Define

In determining what is being evaluated, *systems thinking*, a key principle of developmental evaluation, involves consideration of how micro-,

meso-, and macro-level concerns intersect. Micro-level considerations related to the program itself include how relationships and boundaries between English learning and disciplinary knowledge building are defined and embedded in curriculum and assessments. These then interact with meso-level concerns, such as those related to resourcing and managing the program, including through appropriate staff profiles and training opportunities. Finally, both the micro and meso levels are linked to macro-level concerns, including concerns around the positioning of the program and the university as globally relevant and prestigious, and also to managing the ideological status of English in relation to Bahasa Indonesia and Indonesian national identity.

Adopting a critical applied linguistics perspective, we identified a need to define English proficiency as an essential first step. Our own values were implicated here, as we sought to open for interrogation existing assumptions and beliefs about English as human capital, and to foster, amongst senior leadership and program staff, reflections on the social justice implications of a continued privileging of the English norms enshrined in international English tests, such as the IELTS, TOEFL, and others. We asked our primary stakeholders to consider gathering information about how English as a Lingua Franca (ELF) is used in Indonesia by Indonesians, as a way of re-defining English proficiency for their local context. We argued this would mitigate against a risk of furthering social inequality, since success on international tests can be unduly influenced by wealth (test preparation courses and test taking are extremely costly) in addition to previous access to international education. We also argued that a more localized view of English skills would help to ensure alignment between their curriculum and both the needs of their students and the expectations of employers in the Indonesian context.

Next, we sought to map a tentative theory of change, which would evolve as our understanding emerged of local values and norms around English usage in Indonesia. The *innovation niche* in this context was conceived in terms of both the development of a localized definition of English ability and the development of a theory of change to encompass both language and knowledge building, as set out further below. The program leaders were invested in developing a novel approach to EMI that was localized, relevant, and at the cutting edge in promoting the

status of Indonesian ways of using English and the value of local identities. They could see this as aligning well with principles of social justice in their own discipline area of political science, where they also sought to problematize processes of globalization and the furthering of social inequalities in countries in Asia, including Indonesia.

To support a theoretically robust definition of English ability, we identified literature in the area of ELF, especially research insights into the strategies and practices used by students in ELF settings to negotiate disagreements and shared understandings (e.g. Toomaneejinda & Harding, 2018). In combination, we also drew on insights into the development of interactional competence in a second language, especially since this work has already provided useful insights into how changes in ability might be formally and reliably assessed (Dai & Davey, 2023; Roever & Kasper, 2018). To encapsulate relationships between language practices and knowledge building in academic disciplines, we turned to the field of sociology of education, and specifically Legitimation Code Theory (LCT) (e.g. Maton, 2013, 2014, 2016). LCT provides an innovative theory of knowledge and knowledge building across disciplines, that situates characteristics of language properties, particularly semantics, as central to knowledge transfer; Maton (2013, 2014, 2016) proposes that building capacity for knowledge transfer is contingent on the semantic profile of the written and oral texts (e.g. readings and lectures) to which students are exposed. He conceptualizes a semantic profile in terms of relative movements between abstractness and concreteness in the development of ideas within and across texts, arguing that concept and idea development that follows semantic "wave" profiles (i.e. shifting back and forth between more abstract and more concrete representations of concepts) functions to bridge the gap between more highly decontextualized, 'academic' concepts, and more local, concrete, and context-dependent representations. This wave profile, when produced by students in their own written and oral production, is highly valued as it is viewed as evidence of understanding and an ability to transfer knowledge across contexts. This has been evidenced by research into literacy development outcomes (Freebody, 2013) and student achievement (Matruglio et al., 2013) across different disciplines in school settings. Frost et al. (2020, 2021) have also

demonstrated the utility of LCT principles in designing integrated language assessment tasks (tasks combining comprehension and production skills).

## Frame

Based on the BetterEvaluation Rainbow framework, we planned evaluation questions to gain insights from stakeholders around what was currently happening in the program, including existing strengths and how these can be maintained and further strengthened, and perceived weaknesses. We sought to consider the values and principles of our primary intended users in guiding our conceptualization and definition of program quality, and aimed to understand the culture of the program, the kinds of institutional resourcing available to support innovation in the program, existing feedback mechanisms across different staffing roles and program/institutional leadership, and the profile of program staff in terms of language backgrounds, teaching experience, and expertise areas. Finally, we aimed to understand staff perspectives on areas for improvement in the program, across any aspect they deemed relevant.

Specifically in relation to the *innovation niche*, we also focused on how program stakeholders conceived of relationships between English and content teaching and learning in their discipline area, including how they saw their own roles and responsibilities as educators; we wanted to know if discipline experts were interested in supporting the English language development of students, and if they felt able to do this, as part of developing our understanding of the capacity for innovation in the program. As mentioned earlier, the *utilization focus* of development evaluation emphasizes the need to ensure that key stakeholders are invested in the evaluation process, and that they share a common understanding of program as well as evaluation objectives. In this case, while we knew that program and faculty leaders were invested in promoting improved English proficiency outcomes for students, we did not want to take for granted that teachers would all see themselves as able to ensure this. We also wanted to encourage further thinking about relationships between

English and disciplinary knowledge building, to elaborate our theory of change.

### Describe and Understanding Causes

To this end, we engaged in ongoing focus group meetings with teaching staff to consider and reflect on how they packaged content in their classes, especially in light of the semantic profile model underlying LCT, which we sought to integrate into evaluative thinking across the program. We further conducted interviews with students to gather insights into their experience of learning in the new EMI context, and observations of newly developed EMI classes, to provide ongoing feedback on how these might be adapted to enhance knowledge building opportunities for students, according to our principle of *co-creation*. This fed into an ongoing process of reporting to intended users—the '**describe**' stage of the Rainbow Framework—which we did by preparing brief written summaries and PowerPoint presentations of main findings from focus group meetings and classroom observations, at monthly intervals over the year of the project. The '**understanding causes**' stage of the framework also evolved iteratively, as we sought to articulate interrelationships between language learning and disciplinary knowledge building in relation to an emerging theory of change. Our goal was to embed a set of sustainable evaluation practices into the program, centred around collaboration between disciplinary and academic skills staff, through focus group meetings, and by encouraging peer-to-peer observations of classroom practices. Our aim, ultimately, was to facilitate ongoing innovative and adaptive thinking around program design, which could continue to proceed after the completion of our involvement.

## Discussion

The purpose of this chapter was to set out the pragmatic paradigm and use branch of program evaluation based on Mertens and Wilson (2019), with a focus on the potential for developmental evaluation to inform and

promote innovate design in the context of EMI programs, particularly where these have emerged relatively recently, as is the case in higher education in Indonesia. Our aim was to highlight the value of program evaluation in enabling enhancements to the quality of integrated English and content teaching within disciplines in higher education, especially important given the rapid expansion of EMI and the potential for this to lead to both positive and negative consequences, not only for learning but also for social justice, especially in contexts of high linguistic diversity. We presented EMI in this chapter as a phenomenon which demands an interconnected view of innovation and evaluation processes, as two sides of a coin. We argued that this kind of systems thinking is necessary to enhance program quality where concerns of language are embedded within broader social and policy concerns, and where cause and effect relations cannot be readily articulated, if this is possible at all. We emphasized our ethical position, which is that EMI programs, like all language programs, should seek to innovate in ways that align with social justice principles of inclusion and decolonization, enabling sustainable and culturally meaningful engagement in teaching and learning activities. We illustrated how this might be planned and framed in the context of an emerging EMI program in Indonesia, in the hope that this can inform the efforts of program leaders and teachers in similar contexts in Asia, who are no doubt also seeking to come to terms with the challenges of an expansion of EMI, and who have a vested interest in promoting consideration of its long-term implications and societal impacts. This is taken up further in the following and remaining chapters of the book, as attention turns to an explicit focus on social values and ethics, in the context of constructivist and transformative evaluation approaches.

## References

Basturkmen, H. (2018). Needs analysis and syllabus design for language for specific purposes. In *The encyclopedia of applied linguistics* (pp. 1–8). https://doi.org/10.1002/9781405198431.wbeal0861.pub2

Bolton, K., Bacon-Shone, J., & Botha, W. (2023). Emi (english-medium instruction) across the asian region. *World Englishes, 42*(3), 392-404. https://doi.org/10.1111/weng.12620

Chapelle, C. (2021). Language program evaluation. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. John Wiley & Sons Inc.

Dai, D. W., & Davey, M. (2023). On the promise of using membership categorization analysis to investigate interactional competence. *Applied Linguistics*, amad049. https://doi.org/10.1093/applin/amad049

Dewi, A. (2017). English as a medium of instruction in Indonesian higher education: A study of lecturers' perceptions. In I. B. Fenton-Smith, P. Humphreys, & I. Walkinshaw (Eds.), *English medium instruction in higher education in Asia-Pacific: From policy to pedagogy* (pp. 241–258). Springer.

Elder, C. (2009). Reconciling accountability and development needs in heritage language education: A communication challenge for the evaluation consultant. *Language Teaching Research, 13*, 15–33. https://doi.org/10.1177/1362168808095521

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-based language teaching: Theory and practice*. Cambridge University Press. https://doi.org/10.1017/9781108643689

Fenton-Smith, B., Humphreys, P., & Walkinshaw, I. (Eds.). (2017). *English medium instruction in higher education in Asia-Pacific: From policy to pedagogy*. Springer.

Flowerdew, L. (2012). Needs analysis and curriculum development in ESP. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 325–346). John Wiley & Sons.

Freebody, P. (2013). Knowledge and school talk: Intellectual accommodations to literacy? *Linguistics and Education, 24*, 4–7.

Frost, K., Clothier, J., Huisman, A., & Wigglesworth, G. (2020). Responding to a TOEFL iBT integrated speaking task: Mapping task demands and test-takers' use of stimulus content. *Language Testing, 37*(1), 133–155. https://doi.org/10.1177/0265532219860750

Frost, K., Wigglesworth, G., & Clothier, J. (2021). Relationships between comprehension, strategic behaviors, and content-related aspects of test performances in integrated speaking tasks. *Language Assessment Quarterly, 18*(2), 133–153. https://doi.org/10.1080/15434303.2020.1835918

Hino, N. (2017). The significance of EMI for the learning of EIL in higher education: Four cases from Japan. In B. Fenton-Smith, P. Humphreys, & I. Walkinshaw (Eds.), *English medium instruction in higher education in Asia-Pacific: From policy to pedagogy* (pp. 115–132). Springer.

Hultgren, A. K. (2014). English language use at the internationalised universities of Northern Europe: Is there a correlation between Englishisation and world rank? *Multilingua, 2014*(3/4), 391–414.

Kim, E. G. (2017). English medium instruction in Korean higher education: Challenges and future directions. In B. Fenton-Smith, P. Humphreys, & I. Walkinshaw (Eds.), *English medium instruction in higher education in Asia-Pacific: From policy to pedagogy* (pp. 53–70). Springer.
Kubota, R. (2022). Decolonizing second language writing: Possibilities and challenges. *Journal of Second Language Writing, 58*, 100946.
Lamb, M., Waskita, D., Kuchah, K., Hadisantosa, N., & Ahmad, N.F. (2021). *The state of English as medium of instruction (EMI) in higher education institutions in Indonesia*. British Council.
Lancaster, N. K. (2018). Innovations and Challenges in CLIL Program Evaluation. *Theory Into Practice, 57*(3), 250–257. https://doi.org/10.1080/00405841.2018.1484034.
Larsen-Freeman, D. & Cameron, L. (2008). Complex systems and applied linguistics. .
Lynch, B. (1996). *Language program evaluation*. Cambridge University Press.
Lynch, B. (2003). *Language Assessment and Programme Evaluation*. Edinburgh University Press.
Lynch, B. (2019). *Language assessment and programme evaluation*. Edinburgh University Press. https://doi.org/10.1515/9781474470353
Macaro, E. (2015). English medium instruction: Time to start asking some difficult questions. *Modern English Teacher*, 24(2), 4–8.
Macaro, E. (2022). English medium instruction: What do we know so far and what do we still need to find out? *Language Teaching, 55*, 533–546. https://doi.org/10.1017/S0261444822000052
Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching, 51*(1), 36–76. https://doi.org/10.1017/S0261444817000350
Macaro, E., Hultgren, A., Kirkpatrick, A., & Lasagabaster, D. (2019). English medium instruction: Global views and countries in focus. *Language Teaching, 52*(2), 231–248. https://doi.org/10.1017/S0261444816000380
Maton, K. (2013). Making semantic waves: A key to cumulative knowledge-building. *Linguistics and Education, 24*, 8–22.
Maton, K. (2014). *Knowledge and Knowers. Towards a realist sociology of education*. Routledge.
Maton, K. (2016). Legitimation code theory: Building knowledge about knowledge-building. In K. Maton, S. Hood, & S. Shay (Eds.), *Knowledge-building: Educational studies in Legitimation Code Theory*. Routledge.

Matruglio, E., Maton, K., & Martin, J. (2013). Time travel: The role of temporality in enabling semantic waves in secondary school teaching. *Linguistics and Education, 24*, 38–49.

McKinley, J., & Galloway, N. (Eds.). (2022). *English-medium instruction practices in higher education: International perspectives*. Bloomsbury Publishing.

Mertens, D. M., & Wilson, A. T. (2019). *Program evaluation theory and practice: A comprehensive guide*. Guilford Publications.

Mirhosseini, S.-A. (2018a). Introduction to the special issue: Politics of research in language education. *Critical Inquiry in Language Studies, 15*(4), 231–236. https://doi.org/10.1080/15427587.2018.1442987

Nguyen, H. T. (2022). English-medium instruction in Vietnamese higher education: From government policies to institutional practices. In J. McKinley & N. Galloway (Eds.), *English-medium instruction practices in higher education: International perspectives* (pp. 173–184). Bloomsbury Academic.

Norris, J. M. (2009). Understanding and improving language education through program evaluation: Introduction to the special issue. *Language Teaching Research, 13*(1), 7–13. https://doi.org/10.1177/1362168808095520

Norris, J. M. (2016). Language program evaluation. *The Modern Language Journal, 100*(Suppl), 169–189.

Patton, M. Q. (2008). *Utilization-focused evaluation*. SAGE.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Publications.

Patton, M. Q. (2015a). State of the art and practice of developmental evaluation. In M. Q. Patton, K. McKegg, & N. Wehipeihana (Eds.), *Developmental evaluation exemplars: Principles in practice* (pp. 1–24). Guilford Publications.

Patton, M. Q. (2015b). The developmental evaluation mindset: Eight guiding principles. In M. Q. Patton, K. McKegg, & N. Wehipeihana (Eds.), *Developmental evaluation exemplars: Principles in practice* (pp. 289–312). Guilford Publications.

Phan, L. H. and Doan, B. N. (2020). English as a Medium of Instruction (EMI) in Vietnamese Universities: Policies of Encouragement and Pedagogies of Assumption. In Phan, L. H and Doan, B. N (Eds.), *Higher Education in Market-Oriented Socialist Vietnam, International and Development Education*, https://doi.org/10.1007/978-3-030-46912-2_14

Richards, J. C., & Pun, J. (2021). A typology of English Medium instruction. *RELC Journal*. https://doi.org/10.1177/003368822096858

Roever, C., & Kasper, G. (2018). Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing, 35*(3), 331–355.

Shohamy, E. (2012). A critical perspective on the use of English as a medium of instruction at university. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges* (pp. 196–210). Multilingual Matters.

Smith, G. F., Jung, H., & Zenker, F. (2022). From task-based needs analysis to curriculum evaluation: Putting methodological innovations to the test in an English for academic purposes program. *English for Specific Purposes, 66*, 80–93.

Toomaneejinda, A., & Harding, L. (2018). Disagreement practices in ELF group discussion: Verbal, nonverbal and interactional strategies. *Journal of English as a Lingua Franca, 7*(2), 307–332.

Zein, S., Sukyadi, D., Hamied, F. A., & Lengkanawati, N. S. (2020). English language education in Indonesia: A review of research (2011-2019). *Language Teaching, 53*, 491–523.

Zhao, J., & Dixon, L. Q. (Eds.). (2017). *English-medium instruction in Chinese universities: Perspectives, discourse and evaluation*. Routledge.

**Dr. Kellie Frost** is a Senior Lecturer in ESL and Applied Linguistics at the University of Melbourne. Her research is in the areas of English for academic purposes, language in the workplace, and language testing for immigration and citizenship. She has worked in a range of language program evaluations, with a focus on assessment.